

**IN THE CLAIMS:**

Please replace the following claims for the same-numbered claims in the application.

1. (Currently Amended) A method for clustering data points with defined quantified relationships between them comprising the steps of:
- [[ - ]] obtaining a lead value for each data point, wherein said lead value is derived from any of either by deriving from said quantified relationships or and as given input,
  - [[ - ]] ranking each data point in a lead value sequence list in descending order of lead value,
  - [[ - ]] assigning the a first data point in said lead value sequence list as the a leader of the a first cluster, and
  - [[ - ]] considering each subsequent data point in said lead value sequence list as a leader of a new cluster if its relationship with the leaders of each of the previous clusters is less than a defined threshold value or as a member of one or more clusters at least one cluster where its relationship with the a cluster leader is more than or at least equal to said threshold value[[ - ]],  
and  
generating a text summarization of any of a single document and a collection of documents based on said clustering of data points.
2. (Currently Amended) The method ~~as claimed in~~ of claim 1, wherein said quantified relationships between data points are any of symmetric or and asymmetric quantified relationships.

JP920000447US1

2

3. (Currently Amended) The method ~~as claimed in~~ of claim 1, wherein the lead value of each data point is determined by taking the a sum of relation values of each of the other data points to said data point.
4. (Currently Amended) The method ~~as claimed in~~ of claim 1, wherein said threshold value is adaptively found for a given number of clusters.
5. (Currently Amended) ~~A method for~~ The method of claim 1, further comprising organizing a set of data points into a hierarchy of clusters ~~wherein the method claimed in claim 1 is first used to cluster~~ by clustering the data points into sets of small sizes, wherein each smaller set is further subclustered; ~~using the method and repeatedly~~ repeatedly subclustering said smaller set is repeated until a terminating condition is reached.
6. (Currently Amended) The method ~~as claimed in~~ of claim 1, wherein said step of generating further comprises ~~applied to text summarization of a single document or a collection of documents comprising the steps of:~~
- [[ - ]] segmenting ~~the a~~ given input text into blocks ~~such as~~ comprising sentences, a collection of sentences, and paragraphs,
  - [[ - ]] excluding words belonging to a defined list of ~~'step'~~ defined stop words,
  - [[ - ]] replacing words by their existing unique synonymous word, ~~if it exists,~~ from a given a collection of synonyms,
  - [[ - ]] ~~application of~~ applying stemming algorithms for mapping words to root words,

JP920000447US1

3

[[ - ]] representing the resulting blocks of text, with respect to a dictionary which is either given or computed from the input text, by a binary vector of size equal to the number of words in the dictionary whose  $i$ th element is 1 if an  $i$ th word in the dictionary is present in the block,

[[ - ]] computing the relationship between any data points  $d_i$  and  $d_j$  by evaluating  $R(d_i, d_j) = |d_i \cdot T d_j| / |d_i|$ , wherein  $T$  is a thesaurus matrix whose  $i$ th element reflects ~~the~~ an extent of inclusion of meaning of  $j$ th word in the meaning of  $i$ th word, and

[[ - ]] clustering the data points wherein the lead value of each data point is determined by taking ~~the~~ a sum of relation values of each of the other data points to said data point, wherein the threshold value is adaptively found for a given number of clusters, and ~~the~~ wherein a set of leaders of the resulting clusters summarize ~~the~~ a given text.

7. (Currently Amended) The method ~~as claimed in~~ of claim 6, wherein said dictionary is computed by taking ~~the~~ a fraction of words, excluding the said stop words, with a highest tfidf value, which is given by:

$$\text{tfidf}(w_i) = \text{tfi} * \log(N/\text{dfi}),$$

where  $\text{tfidf}(w_i)$  is the lead value of data point  $w_i$ ,  $\text{tfi}$  = ~~the~~ a number of times the data point  $w_i$  occurred in the a whole text,  $\text{dfi}$  = ~~the~~ a number of documents containing the data point  $w_i$  and  $N$  = ~~the~~ a total number of documents in the text.

8. (Currently Amended) The method ~~as claimed in~~ of claim 6, wherein said thesaurus matrix ~~is either~~ comprises any of a given identity matrix, and a ~~or~~ computed matrix from a collection of documents.

JP920000447US1

4

9. (Currently Amended) The method ~~as claimed in~~ of claim 6, wherein each block is represented by a vector whose *i*th element represents ~~the~~ a frequency of occurrence of said *i*th word in the block.

10. (Currently Amended) ~~A method for~~ The method of claim 6, further comprising organizing a set of text documents into a hierarchy of clusters ~~wherein the method claimed in claim 1 is first used to cluster the~~ by clustering given documents into sets of small sizes, wherein each smaller set is further subclustered; ~~using the method and repeatedly~~ subclustering said smaller set is repeated until a terminating condition is reached.

11. (Currently Amended) The method ~~as claimed in~~ of claim 10, ~~applied to organize the~~ further comprising organizing results returned by ~~any~~ an information retrieval system in response to an user query into an hierarchy of clusters.

12. (Currently Amended) The method ~~as claimed in~~ of claim 11, wherein the hierarchy is used to aid the user in any of modifying ~~his/her~~ a query of said user and ~~and/or in~~ browsing through ~~the~~ said results.

13. (Currently Amended) The method ~~as claimed in~~ of claim 11, wherein the said information retrieval system ~~is any~~ comprises a search engine retrieving Web documents.

A1  
JP920000447US1

5

14. (Currently Amended) The method ~~as claimed in~~ of claim 5, wherein said step of generating is applied to vocabulary organization for a group of documents wherein the data points are the words in the a dictionary of the vocabulary, wherein the lead value of a word is either any of its frequency of occurrence in the collection of documents, the a number of documents containing the word, and a or its tfidf value of said word, wherein the a relationship  $R(d_i, d_j)$  denotes the a fraction of documents containing the a jth word that also contain contains an ith word, and the clustering produced by the application of the method of said data points results in a structured hierarchical organization of the vocabulary.

15. (Currently Amended) The method ~~as claimed in~~ of claim 14, wherein the a structured vocabulary is used to provide text summarization for the associated documents.

16. (Currently Amended) The method ~~as claimed in~~ of claim 14, further comprising applying the clustering applied to customer profiling wherein the a dictionary is built and the vocabulary is organized using the documents that are viewed by the a customer.

17. (Currently Amended) The method ~~as claimed in~~ of claim 5, wherein said data points correspond to the products cataloged in the an electronic store, the lead value of a product is its per unit profit, its per unit value or the a number of items sold per unit time, and the a relationship between the products is either explicitly defined or derived from the purchase data.

18. (Currently Amended) The method ~~as claimed in~~ of claim 17, wherein the a product di is

JP920000447US1

6

related to the a product dj by the a fraction of customer transactions containing dj that also contain di.

19. (Currently Amended) The method ~~as claimed in~~ of claim 17, further comprising applying the clustering to any of applied to analyze an analysis of sales of a store for the a merchant or to organize the, and an organization of a layout of the store to facilitate easy access to products.

20. (Currently Amended) The method ~~as claimed in~~ of claim 17, further comprising applying the clustering applied to personalize the an electronic store layout to an individual customer by using the a relationship that is specific to the individual customer.

21. (Currently Amended) The method ~~as claimed in~~ of claim 5, further comprising applying the clustering applied to customer segmentation for a sales or service organization wherein the data points ~~are the~~ comprise customers in the ~~data base~~ a database, wherein the lead values are ~~their~~ any of a total purchase amount per unit time of said customers, ~~their~~ income of said customers, ~~the~~ a number of times customers visited the an electronic store, ~~or the~~ and a number of items bought by the customer, wherein the a relationship between customers is either explicitly defined or derived from some relevant data, with the a resulting clustering reflecting a structured grouping of customers with similar performances.

22. (Currently Amended) The method ~~as claimed in~~ of claim 21, wherein the a customer di is related to the a customer dj by the a fraction of products bought by dj that are also bought by di.

JP920000447US1

7

23. (Currently Amended) A system for clustering data points with defined quantified relationships between them, said system comprising:

[[ - ]] means for obtaining a lead value for each data point, wherein said lead value is derived from any of either by deriving from said quantified relationships or and as given input,

[[ - ]] means for ranking each data point in a lead value sequence list in descending order of lead value,

[[ - ]] means for assigning the a first data point in said lead value sequence list as the a leader of the a first cluster, and

AI  
[[ - ]] means for considering each subsequent data point in said lead value sequence list as a leader of a new cluster if its relationship with the leaders of each of the previous clusters is less than a defined threshold value or as a member of ~~one or more clusters~~ at least one cluster where its relationship with the a cluster leader is ~~more than or~~ at least equal to said threshold value[[.]], and

means for generating a text summarization of any of a single document and a collection of documents based on said clustering of data points.

24. (Currently Amended) The system ~~as claimed in~~ of claim 23, wherein said quantified relationships between data points are any of symmetric ~~or and~~ asymmetric quantified relationships.

25. (Currently Amended) The system ~~as claimed in~~ of claim 23, wherein the means for

JP920000447US1

obtaining the lead value of each data point is determined by taking ~~the~~ a sum of relation values of each of ~~the~~ other data points to said data point.

26. (Currently Amended) The system ~~as claimed in~~ of claim 23, wherein said threshold value is adaptively found for a given number of clusters.

27. (Currently Amended) The system ~~for~~ of claim 23, further comprising means for organizing a set of data points into a hierarchy of clusters ~~wherein the system claimed in claim 23 is first used to cluster~~ using means for clustering the data points into sets of small sizes, wherein each smaller set is further subclustered; ~~using the method and~~ repeatedly subclustering said smaller set is repeated until a terminating condition is reached.

A/ 28. (Currently Amended) The system ~~as claimed in~~ of claim 23, wherein said means for generating further comprises ~~used for text summarization of a single document or a collection of documents comprising:~~

[[ - ]] means for segmenting ~~the~~ a given input text into blocks ~~such as~~ comprising sentences, a collection of sentences, and paragraphs,

[[ - ]] means for excluding words belonging to a defined list of ~~'stop'~~ defined stop words,

[[ - ]] means for replacing words by their existing unique synonymous word, ~~if it exists,~~ from a given a collection of synonyms,

[[ - ]] means for applying stemming algorithms for mapping words to root words,

[[ - ]] means for representing ~~the~~ resulting blocks of text, with respect to a dictionary which is

JP920000447US1

9



either given or computed from the input text, by a binary vector of size equal to the number of words in the dictionary whose  $i$ th element is 1 if an  $i$ th word in the dictionary is present in the block,

[[ - ]] means for computing the relationship between any data points  $d_i$  and  $d_j$  by evaluating  $R(d_i, d_j) = |d_i \cdot T d_j| / |d_i|$ , wherein  $T$  is a thesaurus matrix whose  $j$ th element reflects the an extent of inclusion of meaning of  $j$ th word in the meaning of  $i$ th word, and

[[ - ]] means for clustering the data points wherein the lead value of each data point is determined by taking the a sum of relation values of each of the other data points to said data point, wherein the threshold value is adaptively found for a given number of clusters, and the wherein a set of leaders of the resulting clusters summarize the a given text.

29. (Currently Amended) The system ~~as claimed in~~ of claim 28, wherein said dictionary is computed by taking the a fraction of words, excluding the said stop words, with a highest tfidf value, which is given by:

$$\text{tfidf}(w_i) = \text{tf}_i * \log(N/d_f),$$

where  $\text{tfidf}(w_i)$  is the lead value of data point  $w_i$ ,  $\text{tf}_i = \text{the a}$  number of times the data point  $w_i$  occurred in the a whole text,  $d_f = \text{the a}$  number of documents containing the data point  $w_i$  and  $N = \text{the a}$  total number of documents in the text.

30. (Currently Amended) The system ~~as claimed in~~ of claim 28, wherein said thesaurus matrix ~~is either~~ comprises any of a given identity matrix, and a or computed matrix from a collection of documents.

31. (Currently Amended) The system ~~as claimed in~~ of claim 28, wherein each block is represented by a vector whose *i*th element represents ~~the~~ a frequency of occurrence of said *i*th word in the block.

32. (Currently Amended) ~~A system for~~ The system of claim 28, further comprising means for organizing a set of text documents into a hierarchy of clusters ~~wherein the system claimed in claim 28 is first used to cluster the~~ by using means for clustering given documents into sets of small sizes, wherein each smaller set is further subclustered; ~~using the method and repeatedly subclustering said smaller set is repeated~~ until a terminating condition is reached.

33. (Currently Amended) The system ~~as claimed in~~ of claim 32, ~~used to organize the further comprising means for organizing~~ results returned by ~~any~~ an information retrieval system in response to an user query into an hierarchy of clusters.

34. (Currently Amended) The system ~~as claimed in~~ of claim 33, wherein the hierarchy is used to aid the user in any of modifying ~~his/her~~ a query of said user and ~~and/or in~~ browsing through the said results.

35. (Currently Amended) The system ~~as claimed in~~ of claim 33, wherein ~~the~~ said information retrieval system ~~is any~~ comprises a search engine retrieving Web documents.

JP920000447US1

11

36. (Currently Amended) The system ~~as claimed in~~ of claim 27, wherein said means for generating is used for vocabulary organization for a group of documents wherein the data points are the words in ~~the~~ a dictionary of the vocabulary, wherein the lead value of a word is ~~either~~ any of its frequency of occurrence in the collection of documents, ~~the~~ a number of documents containing the word, and a ~~or its~~ tfidf value of said word, wherein ~~the~~ a relationship  $R(d_i, d_j)$  denotes ~~the~~ a fraction of documents containing ~~the~~ a  $j$ th word that also ~~contain~~ contains an  $i$ th word, and the clustering ~~produced by the application of the method of said data points~~ results in a structured hierarchical organization of the vocabulary.

37. (Currently Amended) The system ~~as claimed in~~ of claim 36, wherein ~~the~~ a structured vocabulary is used to provide text summarization for ~~the~~ associated documents.

AI  
38. (Currently Amended) The system ~~as claimed in~~ of claim 36, further comprising means for using the clustering used for customer profiling wherein ~~the~~ a dictionary is built and the vocabulary is organized using ~~the~~ documents that are viewed by ~~the~~ a customer.

39. (Currently Amended) The system ~~as claimed in~~ of claim 27, wherein said data points correspond to ~~the~~ products cataloged in ~~the~~ an electronic store, the lead value of a product is its per unit profit, its per unit value or ~~the~~ a number of items sold per unit time, and ~~the~~ a relationship between the products is either explicitly defined or derived from ~~the~~ purchase data.

40. (Currently Amended) The system ~~as claimed in~~ of claim 39, wherein ~~the~~ a product  $d_i$  is

JP920000447US1

12

related to the a product dj by the a fraction of customer transactions containing dj that also contain di.

41. (Currently Amended) The system ~~as claimed in~~ of claim 39, further comprising means for applying the clustering to any of used for analyzing an analysis of sales of a store for the a merchant or for organizing the, and an organization of a layout of the store to facilitate easy access to products.

42. (Currently Amended) The system ~~as claimed in~~ of claim 39, further comprising means for applying the clustering used to personalize the an electronic store layout to an individual customer by using the a relationship that is specific to the individual customer.

AI  
43. (Currently Amended) The system ~~as claimed in~~ of claim 27, further comprising means for applying the clustering used for customer segmentation for a sales or service organization wherein the data points are the comprise customers in the data-base a database, wherein the lead values are their any of a total purchase amount per unit time of said customers, their income of said customers, the a number of times customers visited the an electronic store, or the and a number of items bought by the customer, wherein the a relationship between customers is either explicitly defined or derived from some relevant data, with the a resulting clustering reflecting a structured grouping of customers with similar performances.

44. (Currently Amended) The system ~~as claimed in~~ of claim 43, wherein the a customer di is

JP920000447US1

related to ~~the~~ a customer dj by ~~the~~ a fraction of products bought by dj that are also bought by di.

45. (Currently Amended) A computer program product comprising computer readable program code stored on computer readable storage medium embodied therein for clustering data points with defined quantified relationships between them, comprising:

[[ - ]] computer readable program code means configured for obtaining a lead value for each data point, wherein said lead value is derived from any of either by deriving from said quantified relationships or and as given input,

[[ - ]] computer readable program code means configured for ranking each data point in a lead value sequence list in descending order of lead value,

[[ - ]] computer readable program code means configured for assigning ~~the~~ a first data point in said lead value sequence list as ~~the~~ a leader of ~~the~~ a first cluster, ~~and~~

AI  
[[ - ]] computer readable program code means configured for considering each subsequent data point in said lead value sequence list as a leader of a new cluster if its relationship with ~~the~~ leaders of each of the previous clusters is less than a defined threshold value or as a member of ~~one or more clusters~~ at least one cluster where its relationship with ~~the~~ a cluster leader is ~~more than or at least~~ equal to said threshold value[[.]], and

computer readable program code means configured for generating a text summarization of any of a single document and a collection of documents based on said clustering of data points.

46. (Currently Amended) The computer program product ~~as claimed in~~ of claim 45, wherein

said quantified relationships between data points are any of symmetric ~~or~~ and asymmetric quantified relationships.

47. (Currently Amended) The computer program product ~~as claimed in~~ of claim 45, wherein said computer readable program code means is configured for obtaining the lead value of each data point is determined by taking ~~the~~ a sum of relation values of each of the other data points to said data point.

48. (Currently Amended) The computer program product ~~as claimed in~~ of claim 45, wherein said threshold value is adaptively found for a given number of clusters.

49. (Currently Amended) The computer program product ~~for~~ of claim 45, further comprising computer readable program code means configured for organizing a set of data points into a hierarchy of clusters wherein ~~the computer program product claimed in claim 45 is first used to cluster~~ using computer readable program code means configured for clustering the data points into sets of small sizes, wherein each smaller set is further subclustered; ~~using the method and repeatedly subclustering said smaller set is repeated~~ until a terminating condition is reached.

50. (Currently Amended) The computer program product ~~as claimed in~~ of claim 45, wherein said computer readable program code means configured for generating further comprises ~~used for text summarization of a single document or a collection of documents comprising:~~

[[ - ]] computer readable program code means configured for segmenting ~~the~~ a given input text

JP920000447US1

15

into blocks ~~such as~~ comprising sentences, a collection of sentences, and paragraphs,

[[ - ]] computer readable program code means configured for excluding words belonging to a defined list of ~~'stop'~~ defined stop words,

[[ - ]] computer readable program code means configured for replacing words by their existing unique synonymous word, ~~if it exists,~~ from a given a collection of synonyms,

[[ - ]] computer readable program code means configured for applying stemming algorithms for mapping words to root words,

[[ - ]] computer readable program code means configured for representing ~~the~~ resulting blocks of text, with respect to a dictionary which is either given or computed from the input text, by a binary vector of size equal to the number of words in the dictionary whose *i*th element is 1 if an *i*th word in the dictionary is present in the block,

[[ - ]] computer readable program code means configured for computing the relationship between any data points  $d_i$  and  $d_j$  by evaluating  $R(d_i, d_j) = |d_i \cdot T d_j| / |d_i|$ , wherein  $T$  is a thesaurus matrix whose *j*th element reflects ~~the~~ an extent of inclusion of meaning of *j*th word in the meaning of *i*th word, and

[[ - ]] computer readable program code means configured for clustering the data points wherein the lead value of each data point is determined by taking ~~the~~ a sum of relation values of each of ~~the~~ other data points to said data point, wherein the threshold value is adaptively found for a given number of clusters, and ~~the~~ wherein a set of leaders of ~~the~~ resulting clusters summarize ~~the~~ a given text.

51. (Currently Amended) The computer program product ~~as claimed in~~ of claim 50, wherein

JP920000447US1

16

said dictionary is computed by taking ~~the~~ a fraction of words, excluding ~~the~~ said stop words, with a highest tfidf value, which is given by:

$$\text{tfidf}(w_i) = \text{tf}_i * \log(N/\text{df}_i),$$

where  $\text{tfidf}(w_i)$  is the lead value of data point  $w_i$ ,  $\text{tf}_i = \text{the}$  a number of times the data point  $w_i$  occurred in ~~the~~ a whole text,  $\text{df}_i = \text{the}$  a number of documents containing the data point  $w_i$  and  $N = \text{the}$  a total number of documents in the text.

52. (Currently Amended) The computer program product ~~as claimed in~~ of claim 50, wherein said thesaurus matrix ~~is either~~ comprises any of a given identity matrix, ~~and a~~ or computed matrix from a collection of documents.

53. (Currently Amended) The computer program product ~~as claimed in~~ of claim 50, wherein each block is represented by a vector whose  $i$ th element represents ~~the~~ a frequency of occurrence of said  $i$ th word in the block.

54. (Currently Amended) ~~A computer program product for~~ The computer program product of claim 50, further comprising computer readable program code means configured for organizing a set of text documents into a hierarchy of clusters ~~wherein the computer program product claimed in claim 50 is first used to cluster the~~ by using computer readable program code means configured for clustering given documents into sets of small sizes, wherein each smaller set is further subclustered; ~~using the method and repeatedly~~ subclustering said smaller set is repeated until a terminating condition is reached.

JP920000447US1

17



55. (Currently Amended) The computer program product ~~as claimed in~~ of claim 54, used to ~~organize the~~ further comprising computer readable program code means configured for organizing results returned by ~~any an~~ an information retrieval system in response to an user query into an hierarchy of clusters.

56. (Currently Amended) The computer program product ~~as claimed in~~ of claim 55, wherein the hierarchy is used to aid the user in any of modifying ~~his/her~~ a query of said user and ~~and/or in~~ browsing through the said results.

57. (Currently Amended) The computer program product ~~as claimed in~~ of claim 55, wherein the said information retrieval system ~~is any~~ comprises a search engine retrieving Web documents.

AI  
58. (Currently Amended) The computer program product ~~as claimed in~~ of claim 49, wherein said computer readable program code means configured for generating is used for ~~configured for~~ vocabulary organization for a group of documents wherein the data points are the words in the a dictionary of the vocabulary, wherein the lead value of a word is either any of its frequency of occurrence in the collection of documents, the a number of documents containing the word, and a ~~or its tfidf value of said word~~ wherein the a relationship  $R(d_i, d_j)$  denotes the a fraction of documents containing the a  $j$ th word that also ~~contain~~ contains an  $i$ th word, and the clustering ~~produced by the application of the method~~ of said data points results in a structured hierarchical organization of the vocabulary.

JP920000447US1

18

59. (Currently Amended) The computer program product ~~as claimed in~~ of claim 58, wherein ~~the a~~ structured vocabulary is used to provide text summarization for the associated documents.

60. (Currently Amended) The computer program product ~~as claimed in~~ of claim 58, further comprising computer readable program code means configured for using the clustering ~~configured~~ for customer profiling wherein ~~the a~~ dictionary is built and the vocabulary is organized using the documents that are viewed by ~~the a~~ customer.

AI  
61. (Currently Amended) The computer program product ~~as claimed in~~ of claim 49, wherein said data points correspond to the products cataloged in ~~the an~~ an electronic store, the lead value of a product is its per unit profit, its per unit value or ~~the a~~ number of items sold per unit time, and ~~the a~~ relationship between the products is either explicitly defined or derived from the purchase data.

62. (Currently Amended) The computer program product ~~as claimed in~~ of claim 61, wherein ~~the a~~ product  $d_i$  is related to ~~the a~~ product  $d_j$  by ~~the a~~ fraction of customer transactions containing  $d_j$  that also contain  $d_i$ .

63. (Currently Amended) The computer program product ~~as claimed in~~ of claim 61, further comprising computer readable program code means configured for applying the clustering to any ~~of configured for analyzing an analysis of~~ sales of a store for ~~the a~~ merchant ~~or for organizing~~

JP920000447US1

19

~~the, and an organization of a~~ layout of the store to facilitate easy access to products.

64. (Currently Amended) The computer program product ~~as claimed in~~ of claim 61, further comprising computer readable program code means configured for applying the clustering ~~configured for personalizing to personalize the~~ an electronic store layout to an individual customer by using ~~the a~~ relationship that is specific to the individual customer.

65. (Currently Amended) The computer program product ~~as claimed in~~ of claim 49, further comprising computer readable program code means configured for applying the clustering ~~configured for customer segmentation for a sales or service organization wherein the data points are the~~ comprise customers in the ~~data-base~~ a database, wherein the lead values are their any of a total purchase amount per unit time of said customers, their income of said customers, the a number of times customers visited ~~the an electronic store, or the and a~~ a number of items bought by the customer, wherein the a relationship between customers is either explicitly defined or derived from some relevant data, with ~~the a~~ resulting clustering reflecting a structured grouping of customers with similar performances.

66. (Currently Amended) The computer program product ~~as claimed in~~ of claim 65, wherein the a customer  $d_i$  is related to ~~the a~~ customer  $d_j$  by ~~the a~~ fraction of products bought by  $d_j$  that are also bought by  $d_i$ .